ATTACHMENT G

Strengths and Weaknesses of Artificial Intelligence

The Laboratory for Clinical Decision Making is dedicated to developing new methods by which artificial intelligence (AI) techniques can be applied to computer-based medical diagnosis and treatment. Activities at the laboratory proceed from the premise that a detailed understanding of how expert physicians think will provide a useful model for medical decision-making programs.

The laboratory is an independent member of the national AIM (Artificial Intelligence in Medicine) community and is informally affiliated with SUMEX (Stanford University Medical Experimental Computer). Both groups are devoted to the study, development, and application of AI programs to biomedicine. SUMEX-AIM and the laboratory are supported by the Biotechnology Resources Program of the NIH Division of Research Resources. As part of the Massachusetts Institute of Technology's Laboratory for Computer Science, the Laboratory for Clinical Decision Making has access to powerful computers that provide researchers with the means to build and test programs at MIT. These computers are linked with the distant Stanford computer that serves most other members of SUMEX-AIM. The medical expertise in the Boston area, typified by researchers at Tufts New England Medical Center, Harvard Medical School, and Boston University Medical School, supplies a ready and highly qualified pool of project collaborators.

Since 1978, Dr. Peter Szolovits has directed activities at the laboratory. In an interview with Gregory Freiherr of the Research Resources Information Center, Dr. Szolovits discussed the goals and operation of the laboratory's current projects.

Dr. Peter Szolovits, principal investigator of the Laboratory for Clinical Decision Making.

What is the overall mission of the laboratory itself? Do the projects follow a definite direction?

I've never been a member of the one-goal school of research. Too often people stumble into each other's way when they are all doing the same thing. The intellectual environment is much healthier when people are working on a number of different problems, so I don't have a sense of one burning goal.

The general goal is threefold. First, we are interested in finding ways to improve health care. We want to produce programs that will be put to use, helping doctors improve treatment and therapy. *I think the field has progressed far enough now to challenge* real problems. Second, we examine the development of new artificial intelligence methods for application to biomedical technology. Third, and of particular importance to some of my medical collaborators, we want to get a better understanding of the intellectual tasks associated with practicing medicine. Our collaborators on AI projects teach medicine, and one of them, Stephen Pauker, has a marvelous characterization of how medical students traditionally are trained. He says, "You set them down in a chair, cram them full of facts for 2 years, wheel them around the hospital wards for another couple of years, and hope that by osmosis they pick up whatever it is they need to know." That is a rather crude, but strangely accurate, characterization of medical education. It would certainly be nice if one could explicitly teach the strategies of diagnosis, prognosis, and treatment. These processes aren't taught in medical school today, but if we could understand them better, maybe we *could* teach them.

Will the AI research at the Laboratory for Clinical Decision Making improve health care in the United States?

In 20 years, we are likely to see real payoffs in terms of practical health care. In 10 years, we will be seeing some, but it is difficult to predict exactly what these effects will be. It is also hard to predict the acceptability of the products that we develop, because we don't know if doctors are going to endorse them, nor can we predict all of the legal and ethical questions that may arise about treating people on the basis of decisions made by impersonal machines. If the program makes a mistake, who is responsible? The answers just aren't in. I think those of us in the field today are very hesitant to make definitive statements about what is going to happen.

Where can we expect major applications of Artificial Intelligence to occur?

There are a number of different ways that these programs could be applied, and they lead to slightly different consequences. Probably the first major application will be to build AI programs for education. This bypasses many of the worries that one has about applying AI to clinical medicine. We have worked very hard to make sure that the programs we've built can be used educationally, that they have the capability of making explicit to the user the knowledge involved in the program. The fact that the program has that capability will allow students to run through and challenge decisions made by the computer while getting explanatory materials directly relevant to the cases they are studying. In the long run, AI technology will attack harder medical problems. One kind of situation that we hope to help is that of a physician who faces a difficult problem that he doesn't know how to handle very well. Ordinarily, he would turn to a human consultant for advice, but when the physician is in a remote area, or perhaps an inner-city clinic, the consultant may not be so readily available. We are preparing computer programs that will analyze patient data in various specialties, using the intellectual process of the expert physician. So the primary health impact of the programs we are building will be to help the average physician with difficult cases. In foreign countries, where there is a very severe shortage of physicians, one might be willing to allow paraprofessionals to operate the programs, but in this country the legal questions would be too overwhelming for such an approach.

It has been suggested that certifying AI programs may lessen the complexity of the legal questions that surround their use in clinical situations.

Certifying these programs is a necessary step regardless of the type of application. Dr. Pauker and I have published a set of steps that we believe are important to take in verifying that a program gives good medical advice. Preliminary steps involve retrospective trials conducted with published case histories. Some of our programs have already run through this phase. The next move is to prospective or clinical trials. But despite these trials, no one is going to claim that you can absolutely guarantee or certify that the program is going to make correct decisions all the time, just as you can't certify that a physician is always going to make correct decisions. The best that can be said of a physician is that he has been examined in various ways, he has taken the necessary courses, and he has had the experience. A similar statement might be applied to programs that have passed the various trial phases.

I can see that an AI program could be examined and assessed in much the same way as a physician, but can it be judged as having enough common sense to deal with complex cases?

That is the critical problem. The computer program doesn't have common sense. It knows a lot about a certain specialty area, but it doesn't understand what things in the real world are sensible and what things are not, and that is what worries me. That is why the programs must be used by experienced people, rather than as a sort of automatic health care mechanism. You read science fiction stories of people plugging into their television sets, which are connected to a remote medical computer that tells them how to treat themselves. Well, that is a stupid idea, because a key ingredient is missing. There is nobody to look at the case, no person to say, "This is a hypochondriac," for example. And building into the computer the ability to understand the common sense aspects of the world that would let the computer make judgments extraneous to a medical specialty is the aspect of AI farthest from realization.

In developing AI techniques, are you attempting to duplicate the human mind?

That's rather dramatic terminology, but I think the answer is yes, as concerns high-level programs. It is certainly a less than traditional approach. This is an engineering development, and engineers don't build trucks to walk, for example, so why, my colleagues often ask, are we not trying to build systems better than expert doctors? Doctors are, after all, fallible. The fact is we'd be happy to build programs better than experts, if indeed it were possible. But at present the most promising route is to aim at obtaining the best result that can be achieved through human thought processes. The digitalis program being built here at the resource includes a model of how digitalis is distributed and metabolized in the body, and a model of how the best cardiologists prescribe the drug. We don't claim that the program exactly reflects how doctors think. We simply aim at building a program that accomplishes what the physician tries to do, and the program is built on approximations of how physicians think.

Generally there are two approaches for designing AI programs. One is mathematical, and the other relies on the manipulation of symbols. Are programs at the laboratory designed along these lines?

None of our programs follows any one approach exclusively. Rather, we integrate the two types of computing, using each one for tasks to which it is best suited. In order to express a symptom, for example, it is necessary to use symbols. I could not express that a patient's toe hurts by using differential equations. But without probabilistic or mathematical estimates, there is no quick way to differentiate between any of the hundreds of different possible causes. The patient may have stubbed his toe, or he may be suffering from some bizarre disease. A program without mathematical computation cannot reject either of these possibilities or the many others between these two extremes until there is definitive evidence. A program of this type is very inefficient.

So symbolic programs typically bring some kind of probabilistic evidence to bear, and they may be sort of symbolic probabilities, like *this is very likely* or *this is very unlikely*. In fact, if you look at all the leading AI medicine programs in this country, every one of them has a little number cruncher that checks each hypothesis as it arises to see if it is worth pursuing, and if so, this number cruncher is then used to rank hypotheses according to their probability of being correct.

What happens when the program encounters contradictory information?

It tends to resolve inconsistencies by numerical methods, which is a weakness. When physicians dealing with difficult cases get contradictory information, they realize that there is something more complicated going on. Contradictions tell the physician to go deeper, to think harder. AI programs, on the other hand, get the answer by subtracting the negative evidence from the positive. So currently we are trying to learn how the expert physician reaches these deeper levels of understanding.

SZOLOVITZ *continued*

It seems that physicians rely on common sense to deal with these problems.

I don't think it's the doctor's common sense that takes over, but his expertise. We have evidence that the more expert the doctor is, the better he is at solving problems. For example, if you present a medical student with a difficult situation, very often he will throw up his hands and say, "Good God, I don't know." But if you take a doctor who has been practicing for a long time, and present him with the same problem, he is very likely to say, "Oh yes, there is this set of complications that comes up occasionally, and if that is indeed the case, then I can explain that factor." This is expertise, not common sense.

Building a program that contains in-depth knowledge and decision-making skills requires an enormous amount of time and effort. Isn't there a simpler method for using the computer to diagnose diseases?

There are non-AI approaches to diagnosis. For example, the decision-algorithm approach to medical decision-making is still relatively popular. In fact, there are a number of decision algorithms already in clinical use. One is used to diagnose urinary tract infections, for example. These algorithms include instructions such as, "If the patient answers 'yes' to more than three of the questions that address various symptoms of a disease, call the doctor immediately." They are simply flowcharts, basically separating the healthy from the sick. And these algorithms are pretty successful, but they're very limited in scope. They don't go very deep into medical considerations.

People have tried extending the flowchart approach to much more complex areas of medical application. For example, a flowchart program in the acid-base electrolyte domain has been built. It began 8 or 10 years ago as a concise categorization of all the different acid-base problems that were understood, and then it was expanded as the developer found more and more cases that didn't fit in the original conception. And as new information became available, the program kept growing. At present, it is about 150 pages written in a rather dense programming language. It's difficult to update, and it's all but impossible to certify as doing what it's supposed to do. The program has knowledge, but it has no conception of the knowledge.

How do you go about giving the program a conception of what its knowledge is?

In the AI representational approaches, one tries to build computer languages that efficiently represent knowledge. Bill Martin, who preceded me as principal investigator, has been directing research toward developing a language that may allow us to program very large volumes of real-world knowledge. When finished, anything you can say in English you will be able to say in this language. At the same time, the language will logically tie together the various parts of the reasoning process. So when you are searching within a data base for some relevant fact, you won't have to search the whole data base. It is indexed appropriately so that you can find what you are looking for. In fact, the explanation capabilities that we've built into the digitalis program exploit these knowledge representations, so that decisions can be explained in English rather than in a hard-to-understand programming language. This is one of the big advantages of AI.

Rather than providing the computer with knowledge concerning a specific domain, couldn't the computer be programmed to simulate specific organ systems, so that when biological malfunctions occur, the computer could be used to determine their exact cause?

It would seem to be a good way—if not the best way—to get the answers. The problem is that simulations of this type do not work. Some years ago, Bill Schwartz and one of his medical colleagues tried to simulate the function of the kidneys, and they found that you start getting into physiological phenomona that nobody understands. For instance, although you think of the fluid flow through the nephron as uniform, it isn't. In order to build an accurate simulation, you have to understand many things that are currently far beyond our grasp.

It has been said that the Boston area has one of the highest concentrations of medical experts in the country. Has this helped you?

One of the hardest things in applying AI to medicine is getting a collaborative group that works well together. Much depends on personal chemistry and accidents of who you happen to meet and how you get involved with people. So, yes, being in the area is a great help. We spend a fair amount of time down at Tufts New England Medical Center or one of the other hospitals, because when either a researcher or a student gets involved in the project, it is important to develop a feel for what that area of medicine is like. You don't get that from reading medical books. You get it from tagging along on rounds and going to morning report, conferences, and meetings. Often, when we bring a group of new people into a project at one time, we have a tutorial session where we hire some fellows, usually from Tufts, to teach us something about whatever domain in medicine it is that we are interested in.

Despite the advances made in attracting collaborators and the success of several AI programs, many scientists are skeptical that AI will ever be practical. How do you get around this attitude?

I don't think you can get around it completely. Sure, as AI programs get better the number of people who say AI won't work will decrease, and that will make it easier to apply these programs in their target areas. But there will always be skeptics, people who change the definition of intelligence so that it's more elaborate than the ability of the best program. There's a joke among AI researchers that as soon as a program works, it's no longer AI.

Do you subscribe to that idea?

No. But you'd be surprised how many do. ■